Les outils de la veille pédagogique

Jean-Paul Pinte, enseignant-chercheur en veille
Université Catholique de Lille, FRANCE
pintejp@aol.com

Réflexion pédagogique

Résumé

Parmi les quelques questions vives qui animent les travaux de recherche pour le développement des TIC en éducation se pose aujourd'hui essentiellement celle de la description et de la recherche pertinente de ressources pédagogiques au travers des réseaux d'information numérisés.

En effet, enseignants comme étudiants sont globalement à la phase où l'ampleur quantitative des ressources sur le réseau a pour rançon une pauvreté qualitative du traitement de « l'interprétation » de ces données. Le concept de veille pédagogique, décrit dans cette communication, fournit une des réponses à cette question. Les outils de recherche actuels et en cours de développement qui y sont présentés ont pour principal objet la recherche d'information dynamique liée à des connaissances mises en forme dans une structure taxinomique interactive.

Abstract

Among the pressing questions that today provide impetus for research into the development of ITCs in education is essentially that of the description and search for relevant educational resources through digital information networks. Indeed, students and teachers alike have generally reached the point where the quantitative aspect of network resources has a counterpart in the qualitative poverty of data interpretation and treatment.

The concept of "education watch" described in this text supplies one answer to this question. The research tools, both existing and in development, presented here primarily focus on searching for dynamic information linked to the knowledge formulated in an interactive taxonomic structure.

À la rentrée 2005, un nombre encore plus important d'étudiants arrivera en cours avec comme outil complémentaire d'accès au savoir « la grande toile ». Dans les classes de la maternelle et du primaire, ils seront nombreux au cours des prochaines années à n'avoir connu d'autre monde que celui du Web. En effet, l'acquisition du réseau Internet, confortée par l'accès haut débit, s'avère aujourd'hui de plus en plus rapide avec les incitations au divertissement, les recherches éducatives, les échanges avec les amis et l'écoute de la musique facilités par cet outil formidable à l'école et à la maison.

Pour l'apprenant qui effectue des tâches, les nouvelles technologies de l'information et de la communication (NTIC) constituent de plus en plus un atout qui le place dans une relation nouvelle par rapport à l'enseignant.

Le potentiel de ces outils est surtout lié à leur flexibilité et à leur capacité de véhiculer l'information de façon multimodale. Les TIC sont aussi liées à la fascination qu'elles exercent, surtout à l'époque de l'adolescence alors que les représenta-

tions négatives ou faussées de soi-même et de son travail sont fréquentes.

Les étudiants qui ont des difficultés à suivre une leçon magistrale monomodale, qui repose tout entière sur l'écoute de la langue orale, peuvent modifier la représentation négative qu'ils ont de leurs capacités à partir du moment où ils suivent la même leçon, au moyen de fiches ou d'hypertextes, sur leur écran.

Mais le Web leur permet aussi de trouver des travaux faits pour eux, sur mesure, qu'ils pourront « copier-coller » en deux secondes et ensuite présenter à leur enseignant comme étant un original. Pendant que les jeunes découvrent Internet (et ils sont très rapides!), les enseignants, eux, doivent commencer à se méfier de cette montée en charge de petits futés qui se servent du réseau et trouvent l'information par simple « sérendipité »[1] dans ce monde d'infobésité[2], et s'intégrer peu à peu dans l'accompagnement à la recherche de l'information pertinente et à son exploitation dans le cadre des cours, qu'ils soient ou non à distance.

Heureusement, il existe aujourd'hui des outils qui peuvent aider les apprenants comme les enseignants à faire face à ces nouvelles tâches et à ces nouveaux défis. Il s'agit des outils de veille pédagogique, l'objet du présent article.

1. Passage des TIC à d'autres TIC

Les TIC sont perçues comme un défi en ce début de siècle, dans la mesure où elles sont vouées à mettre en crise et à désarticuler le système éducatif. Elles constituent aussi un véritable pari puisqu'elles peuvent contribuer à modifier l'approche de fond des acteurs éducatifs en ce qui concerne l'autonomie des apprenants.

Alors qu'elles ont acquis une position prééminente dans la société, les TIC tardent cependant à trouver leur place réelle dans le système éducatif et, alors que les jeunes sont en pleine immersion dans ce que l'on pourrait appeler « les multimédias », les instances éducatives tendent encore à préserver le seul mode de l'écriture/lecture.

Les dernières avancées en ce qui concerne la mise en place de plates-formes de ressources en ligne et de formation à distance orientent inévitablement nos systèmes éducatifs vers les auto-apprentissages.

Deux problèmes freinent aujourd'hui le développement des TIC :

1. La production actuelle des ressources n'est pas guidée par des critères didactiques et pédagogiques, mais par des critères liés aux intérêts techniques et au profit, comme le précise P. Moeglin (1996) en observant que les produits pédagogiques ne sont pas conçus dans une perspective pédagogique et qu'ils exigent en fait de l'apprenant qu'il crée lui-même une convergence dans un produit qui n'en a pas (références souvent à des modèles pédagogiques sous-jacents contradictoires, taylorisme et cognitivisme), bref que l'apprenant soit autonome *a priori*[3].

2. Le second problème, constaté surtout en Europe, est lié à l'attitude des enseignants envers les TIC. Les enseignants ont des difficultés à comprendre le caractère structurel des changements qu'on leur demande de gérer et d'accompagner. Le problème auquel les systèmes éducatifs doivent se mesurer est celui de la vitesse des changements extra-muros, qui semble incompatible avec la durée nécessaire que demande un changement culturel et professionnel significatif.

Il faut aussi souligner le décalage qui existe entre les savoirs académiques transmis par les enseignants et ce que les étudiants en perçoivent aujourd'hui avec l'avènement des NTIC, et définir les éléments favorisant une interaction entre l'enseignant et l'étudiant. Pour cette approche, il est nécessaire de prendre en compte les conceptions de l'étudiant et de convenir que celui-ci réfléchisse sur celles de l'enseignant, d'où un processus d'apprentissage conçu comme un dialogue entre ces deux parties.

Des TIC (technologies de l'information et de la communication), il nous faut aujourd'hui passer aux TIC (technologies pour l'intelligence et la connaissance), et la veille pédagogique peut accompagner ce changement.

2. De l'utilité d'une veille pédagogique

Avec l'avènement des TIC, les conditions dans lesquelles la connaissance est acquise, conservée et produite sont totalement modifiées, et l'économie des médias de masse est remaniée en profondeur. Aussi, quel que soit le public utilisateur dans notre société, de nouveaux comportements au regard de la connaissance apparaissent. On assiste à l'émergence d'un nouveau paradigme, celui de la communication dont la dynamique est présente dans les entreprises et les institutions, et plus largement sur le plan politique mondial. Désormais, la gestion de la connaissance s'inscrit naturellement dans la sphère des activités humaines. Elle aussi est confrontée aux mutations qu'engendre l'effervescence de la communication dans nos processus de travail. Ces mutations remodèlent aussi la fonction documentaire et dessinent ses futures caractéristiques.

La vie intellectuelle est aussi marquée par une médiatisation à distance (télémédiatisation) grandissante d'une information

qui vient de loin, est reprise ici, pour être transformée là et inscrite dans un autre contexte ailleurs encore. La télémédiatisation des connaissances est un enjeu majeur pour les activités documentaires du futur. La sphère de la communication étendant son emprise dans ce domaine par le biais des réseaux, l'activité de gestion des connaissances leur est peu ou prou rattachée depuis assez récemment et se trouve encore complexifiée par l'incursion de nouveaux vecteurs de connaissance avec Internet.

Si on la compare à celle de la recherche documentaire, la connaissance que nous produisons aujourd'hui dans la recherche d'information électronique est davantage le fruit d'échanges et de renouvellements opérés dans l'hypertextualité, la numérisation, la virtualité et la délocalisation affranchies des contraintes physiques (donc de l'exemplarisation, du stockage et de la description physique). La reproductibilité et son aptitude à être manipulée à l'aide, par exemple, des modifications de format et du mode d'affichage, sont autant de caractéristiques propres au document numérique.

Les modes de recherche se voient eux aussi transformés en passant du séquentiel aux combinés grâce aux opérateurs booléens et mots-clés d'abord, puis aux paramètres avec les différents opérateurs rendus nécessaires par la recherche en texte intégral. Il en résulte un intérêt pour une veille informative et technologique sur les réseaux de la part de tous les acteurs de l'éducation et la nécessité de former les apprenants à la réflexion préalable à tout acte de recherche ainsi qu'à la vigilance qu'exige, après la collecte, le tri des résultats. Dans un contexte de bruit informationnel (voir le Schéma 1), les réponses des systèmes sont nombreuses et approximatives. Il leur faudra affiner leurs résultats *a posteriori* et les rendre aptes à analyser et évaluer l'intérêt des documents retournés, et à utiliser des techniques comme celles du panier, de la recherche dans le document, voire de la fouille des corpus rapatriés sur les machines en local.

Ainsi, aujourd'hui, pour un enseignant comme pour un apprenant, la recherche documentaire s'est transformée en recherche d'information dont il est bon de rappeler les trois étapes :

1. collecte d'information (indexation de la question, mobilisation des idées, reformulation éventuelle);
2. traitement de l'information (sélection de l'information, traitement de l'information avec prise de notes et synthèse);
3. communication (restitution / production et évaluation).

Malheureusement, ce que ne nous indique pas cette liste, ce sont les interactions entre les différentes étapes. Ainsi, la collecte d'information est directement orientée par le type de restitution prescrit pour le travail engagé, l'évaluation est nécessaire à toutes les étapes et la sélection des informations dans les documents requiert les mêmes compétences que celles de la démarche tout entière (recherche des éléments pertinents en fonction des consignes de départ, évaluation du lot d'information, fin ou reprise éventuelle du processus de recherche).

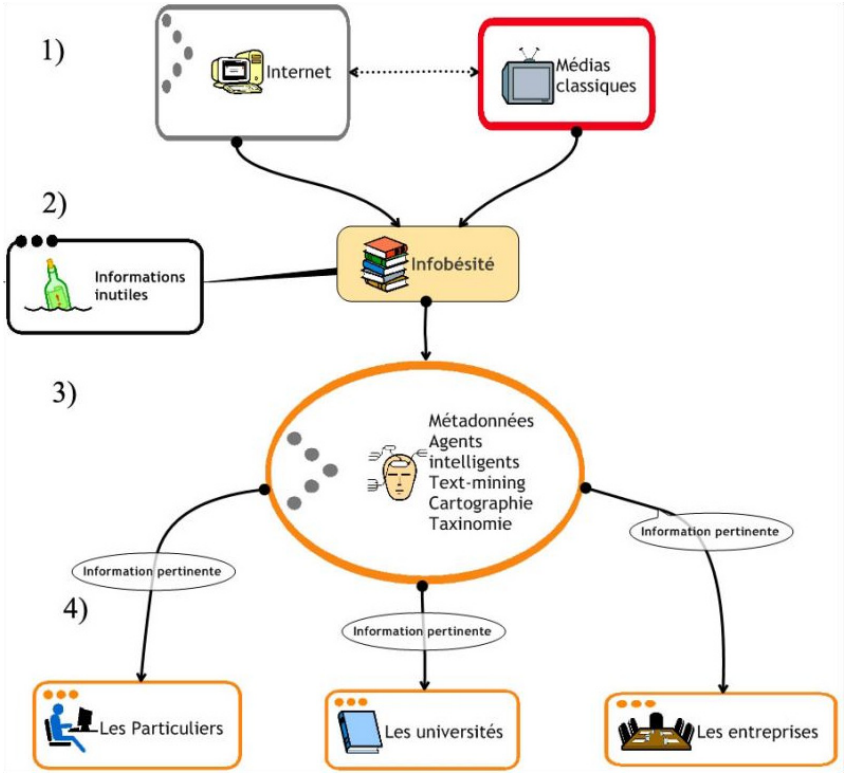
C'est dans cette démarche complexe prenant en compte de nombreux paramètres (sources à ma disposition, procédures d'accès à l'information, consignes données, mots-clés judicieux, etc.) et s'accompagnant d'un esprit d'analyse constamment en éveil que se fondent les prémices du concept de veille pédagogique.

Cette démarche intellectuelle trouve sa particularité dans une dynamique décomposable en quatre temps :

1. formulation d'hypothèses de recherche;
2. vérification scrupuleuse des pistes de recherche;
3. évaluation des résultats;
4. prise de recul, réorientation de la recherche, reformulation de la requête dans la plupart des cas.

Ce processus intellectuel s'appuie sur la mémoire du long terme (culture générale, culture documentaire), sur la mémoire du court terme (qu'est-ce que je viens tout juste de trouver?), sur les possibilités d'abstraction (quel est le mot-clé générique de ma recherche?), sur une connaissance fine du langage et sur un esprit critique, curieux, méthodique et rigoureux avec une once d'intuition. Il s'apparente donc plus à la résolution de problème et on comprendra qu'il soit spontanément adopté par les autodidactes.

L'enseignement de ce que l'on peut appeler maintenant une méthode de veille, complétée et adaptée par les précautions que prennent les professionnels de l'information tels les journalistes et les responsables des communications des entreprises (vérification des sources, croisement des informations, distinction entre les faits et leur commentaire, etc.), permet de diriger nos étudiants vers cette société de l'information, caractérisée par la multiplicité des sources auxquelles on accède en temps réel, par la profusion d'informations non contrôlées, redondantes, obsolètes et souvent orientées, comme le démontre actuellement le développement du concept de « googlelisation » de l'information qui nous apprend que 72 % des recherches effectuées sur le Web le sont à partir de Google (Lagane, 2004).



1- Les principales sources d'information viennent de l'ensemble des médias, c'est-à-dire de la télé, de la radio, des journaux, etc., et bien sûr d'Internet. Ces deux grandes sources d'information se complètent souvent entre elles aujourd'hui par le jeu de l'hypertexte et donc le renvoi vers une ressource de la toile.

2- Face à cette massification de l'information, l'homme arrive vite à une surcharge informationnelle, qualifiée aujourd'hui d'infobésité. De cette masse d'information, il lui faut alors sélectionner et extraire l'essentiel.

3- Il existe différents moyens et méthodes pour sérier l'information, comme l'utilisation de métadonnées (données sur des données) pour décrire le contenu et la fonctionnalité de sources, les agents intelligents pour automatiser et filtrer l'information, l'exploration de texte (*text mining*) ou l'analyse de texte pour en retirer l'essentiel, la cartographie pour représenter de façon claire et synthétique de l'information sous forme de carte et, enfin, la taxinomie pour classer et répertorier l'information afin d'aller directement à l'essentiel.

4- Ces méthodes vont permettre d'extraire de la masse l'information pertinente pour toute personne qui recherche sur Internet; pour l'université dans le cadre d'une gestion de la connaissance; et pour l'entreprise, afin de réaliser des gains de temps dans la recherche d'une information stratégique dont elle a besoin (sur un produit, sur un concurrent, sur un marché, etc.).

Schéma 1. L'homme face à l'infobésité (Pinte, juin 2003)

3. Internet et les multimédias au cœur du processus de veille pédagogique

Il y a seulement quelques années encore, une personne en quête d'information savait qu'il lui fallait se rendre dans un lieu documentaire pour satisfaire sa demande après avoir exploité la richesse du lieu. Dans ce contexte, il lui fallait alors comprendre au minimum les techniques professionnelles conçues par des experts de l'information et de la documentation et se les approprier pour retrouver le rangement ou le classement d'une ressource, d'un article ou d'un livre, ce qui n'était pas toujours une chose facile à l'époque.

Les années 1980 ont fondé de grands espoirs sur la recherche documentaire informatisée avec l'arrivée des ordinateurs et la possibilité de formuler plus ouvertement les recherches, et avec la diversification des accès au signalement des ressources, qui ont sensiblement modifié la relation de l'usager au document.

Depuis 1993, le réseau des réseaux a entraîné le concept de cyberespace qualifié de « nouvel emblème de l'encyclopédisme universel » par Morizio (2002). Son introduction dans les centres de ressources documentaires et à la maison a rapidement mis en évidence les raccourcis entre accès à l'information et construction de savoirs.

Quelques chiffres peuvent à cet effet nous donner aujourd'hui le vertige :

- Le cabinet eTForecasts estime que le nombre d'internautes dépassera le milliard d'ici la fin de 2005. (Voir : http://www.etforecasts.com)
- Selon International Data Corporation, plus de 30 milliards de courriels seront échangés par jour à la fin de 2005. (Voir : http://www.idc.com)

- Online Computer Library Center, Inc. évalue à 9,04 millions le nombre de sites uniques sur le Web. (Voir : http://wcp.oclc.org)
- La société Cyveillance estime que le Web visible contient aujourd'hui plus de 8 milliards de pages. (Voir : http://www.cyveillance.com)
- Selon la société BrightPlanet, le Web invisible[4] contiendrait 550 milliards de documents à l'heure où cet article est rédigé. (Voir http://www.press.umich.edu/jep/07-01/bergman.html)

On comprend aisément que devant un tel gisement informationnel, tout acteur du système éducatif soit si souvent insatisfait du résultat de ses recherches sur le Net.

Aussi, en entrant dans ce nouveau millénaire, le milieu éducatif s'interroge aujourd'hui plus particulièrement sur les relations entre la recherche documentaire, l'utilisation des nouvelles technologies et les apprentissages. Les différents modes de structuration des connaissances dans les documents numériques deviennent aussi des centres d'intérêt pour les chercheurs en psychologie cognitive.

Internet et les médias classiques doivent donc être perçus aujourd'hui par les enseignants et par les étudiants comme des outils leur permettant de soutenir les préparations de cours, de trouver des outils pédagogiques concrets, d'effectuer des recherches sur une problématique ou un thème donné et, ainsi, d'actualiser l'enseignement en fournissant des références additionnelles aux cours dispensés en face à face.

La recherche d'information ne doit pas être uniquement une affaire de machine, mais une affaire d'« homme », produisant du savoir, le mettant en scène, le conservant, le communiquant et se l'appropriant pour être et pour agir.

Depuis la mise en place d'outils d'apprentissage et de plates-formes de ressources pédagogiques en ligne, on voit poindre ici et là l'évolution de plusieurs paramètres :

- Les centres de ressources permettent de développer le travail coopératif et garantissent à l'apprenant non plus de se fondre dans le groupe, mais de se construire grâce à des relations avec autrui et de développer sa personnalité globale et sa relation à l'altérité.
- L'apprenant acquiert des compétences d'apprentissage transférables qui lui assurent la possibilité de travailler en autonomie le moment venu.
- L'intégration des multimédias offre, de plus, une gamme d'exploitations individuelles variées où l'enseignant se doit de préparer des tâches ou des scénarios pédagogiques.
- L'apprenant, au travers de l'intégration des multimédias dans l'apprentissage, a la liberté de choisir ses parcours, d'effectuer des découvertes. L'enseignant est là pour écouter, encourager et permettre aux apprenants de confronter leurs observations et leurs découvertes.

La transition vers l'accès aux nouveaux dispositifs cités ci-dessus doit se préparer par des étapes où l'apprenant apprend à utiliser les multimédias en bénéficiant des conseils de l'enseignant.

> « Peut-être devrions-nous utiliser les technologies pour fournir une base de données riche en information et démonstrations. Fournir un laboratoire d'apprentissage dans lequel les étudiants font de l'exploration et résolvent des problèmes qui sont proposés par les enseignants. »

De cette façon, selon Norman (1991), les professeurs deviennent des assistants dans la découverte des connaissances, des guides dans l'exploration, la structuration de la compréhension de l'étudiant.

La mission de la veille pédagogique vise dans ce cadre à promouvoir et à soutenir l'intégration des TIC dans l'enseignement, et cette mission peut se traduire par trois axes fondamentaux :

- informer les intervenants de l'éducation sur les ressources didactiques disponibles et les pratiques pédagogiques innovantes en matière de TIC;
- appuyer les acteurs de l'enseignement dans leur démarche d'appropriation pédagogique des TIC par le développement d'habiletés telles que la recherche efficace de ressources disciplinaires, voire la création de scénarios pédagogiques intégrant les TIC;
- traquer, traiter, partager et capitaliser l'information pour une meilleure gestion des connaissances.

À ces effets, tous les acteurs de l'éducation, de l'apprenant à l'enseignant en passant par l'organisation, se devraient de se constituer une boîte à outils ou mallette de veille pédagogique qui va des compétences relatives à l'analyse du contexte cyberculturel, au décodage de la communication et à la structuration de la relation pédagogique avec les TIC.

4. Outils de la veille pédagogique

Le filtrage d'information en ligne, plus communément appelé de nos jours « recherche d'information en ligne » grâce à des moteurs de recherche et à des logiciels documentaires, n'est pas nouveau et remonte à une quarantaine d'années. Comme le signale Perriault (2002, p. 64), l'histoire de cette technique est intéressante car, après avoir tendu vers le « tout automatique », on est en train de réinsérer l'intervention humaine dans son dispositif.

Sans entrer dans l'historique complet de la recherche automatisée de documents, il convient néanmoins de signaler les travaux qui ont suivi ceux de Paul Otlet en 1904 avec sa classification décimale universelle (UDC Consortium, n. d.). Dès 1960, les travaux de Grolier, et notamment de Gardin avec son équipe du CNRS (Cros, Gardin et Lévy, 1964), réalisent Syntol (*Syntagmatic Organisational Language*). Ce langage formalisé permettait en effet de construire des requêtes[5] très fouillées pour consulter une base de données contenant des documents indexés avec la même finesse. Cette finesse, qui n'est plus de mise aujourd'hui, résidait dans la représentation de la question posée, d'une part, à l'aide de descripteurs ou mots-clés, que l'on utilise encore aujourd'hui dans les moteurs de recherche et, d'autre part, dans les relations sémantiques qui indiquaient le rapport de sens entre les descripteurs « consécutif », « associatif », « prédicatif », etc. (on pouvait distinguer une question concernant le chien de mon père de celle concernant le père de mon chien, pour reprendre l'exemple de Perriault).

Notons que les moteurs actuels n'utilisent plus cette fonctionnalité et se contentent de ET, de OU et de SAUF, que l'on peut interposer entre des termes ou des groupes de termes encadrés par des parenthèses.

À tout cela s'ajoute la volonté des acteurs d'accéder à des informations de plus en plus pertinentes en caressant de près l'idée de les obtenir si possible directement dans leur courriel sans en faire la demande…

Devant toutes ces attentes, l'évolution d'Internet et le déluge informationnel engendré, cette analyse a maintenant pour objectif de fournir les bases nécessaires et suffisantes à tous les acteurs de l'éducation pour trouver de manière efficace l'information sur le réseau. Une première partie abordera une typologie des informations présentes sur le Web tout en posant la faisabilité de leur collecte, une seconde sera consacrée aux grandes familles d'outils manuels et automatisés de recherche. Pour conclure, l'auteur propose de consulter en ligne sur Internet une liste synthétique de liens et de ressources utiles, voire indispensables, sélectionnés à votre intention pour vous aider à mieux appréhender le concept de veille pédagogique.

4.1 De l'information blanche à l'information noire

La définition d'une typologie de l'information est un préalable à toute activité de collecte et de traitement dans le processus de veille. Celle qui est présentée dans le Tableau 1 ci-dessous reprend les travaux de divers auteurs (Dou, 1995; Achard et Bernat, 1998[6]).

L'activité de veille pédagogique repose sur le concept d'autonomie informationnelle. Cela suppose de définir par avance la base à partir de laquelle sera développé le processus, autrement dit : quel type d'information entre dans cette problématique? À quel niveau décisionnel peut-elle être utilisée? Quel est son domaine opératoire?

Tableau 1. Typologie de l'information (selon Franck Bulinge)

Information	Blanche	Grise	Noire
Type	Scientifique, technologique, commercial, juridique, financier, stratégique, personnel		
Niveau	Tactique, opérationnel, stratégique		
Domaine opératoire	Documentaire, de situation, d'alerte		
Intérêt	Fatal, utile, pertinent	Pertinent, critique	Critique
Accès	Public	Restreint	Strictement limité
Classification	Non protégée	Protection restreinte	Confidentielle - Secrète
Disponibilité	80 %	15 %	5 %
Acquisition -Exploitation	Légale sous réserve de respecter les droits de propriété.	Domaine juridique non clairement défini. Risques d'ordre jurisprudentiel.	Illégale, l'acquisition relève de l'espionnage. Risques très élevés.
Forme	Formelle (texte) ou informelle (conversation, rumeur)		
Sources	Ouvertes	Autorisées - Fermées	Clandestines
Coût	Faible	Faible	Élevé
Rentabilité	Élevée	Très élevée	Faible

Tableau 2. Sources et supports d'information (selon Franck Bulinge)

Information (Support)	Structurée	Non structurée
Documentaire	Rapports d'étonnement Enquêtes Formulaires Questionnaires Annuaires téléphoniques Tarifs Normes	Presse (journaux, revues) Ouvrages, livres Rapports d'ambassade (PEE) Publications des cabinets spécialisés Documents légaux (bilans, rapports annuels) Courrier Journaux d'entreprise, tracts Rapports de stages, thèses Plaquettes commerciales Manuels d'utilisation Revues et documents en ligne
Électronique	Bases de données Brevets	Sites Internet Forums de discussion Listes de diffusion
Multimédia		Enregistrements sonores Films documentaires Reportages, photographies
Relationnel		Expertise interne Échanges clients / fournisseurs Colloques, séminaires, foires Réseaux et chambres consulaires, clubs d'entreprise
Informel		Conversations, indiscrétions, rumeurs, *e-mails*

En quelque sorte, quel niveau d'attention temporelle requiert-elle? Quels sont son intérêt, son accès, sa forme, ses sources, etc.?

Dans ce tableau, seules l'information blanche et l'information grise concernent la veille; la troisième, l'information secrète, relève de l'activité des services spéciaux comme ceux de la défense d'un pays.

Le principe essentiel de la veille menée par un étudiant ou un enseignant est fondé sur le recueil de l'information blanche, l'information grise marquant une zone de non droit séparant d'une manière relativement floue les territoires autorisés des territoires interdits.

Globalement, nous pouvons définir les sources d'information du point de vue de la structuration, puis du point de vue des supports, comme présenté dans le Schéma 3 ci-dessous.

Dans sa thèse de doctorat, Franck Bulinge (2002, page 184) définit l'information structurée comme une information textuelle dont le traitement peut être assuré automatiquement et en totalité par des outils informatiques (ex. : des bases de données relationnelles).

L'analyse du Tableau 2 nous permet de constater que l'information non structurée est beaucoup plus abondante et diversifiée que l'information structurée.

La problématique de la veille est donc complexe : il s'agit de collecter et d'exploiter de l'information ouverte dont la diversité est telle qu'il est impossible d'envisager un système de traitement homogène.

À cet effet, il convient, avant de se lancer dans l'utilisation d'outils de veille automatisés, de bien connaître les familles d'outils auxquelles l'acteur de l'éducation a accès pour aboutir dans sa demande d'information.

4.2 Les familles d'outils

- Les répertoires ou les annuaires

Ils sont parmi les premiers outils de recherche à avoir fait leur apparition sur le Web, le plus connu, Yahoo, ayant été créé en 1994 par David Filo et Jerry Yang, deux étudiants de l'Université de Standford aux États-Unis.

Le but des annuaires est de répertorier les sites Web et de les classer dans des catégories thématiques pour faciliter leur identification par l'internaute. À partir d'un

thème, une série de sous-thèmes est proposée à l'utilisateur pour aboutir enfin à une liste de sites répondant à sa question.

On peut distinguer trois types d'annuaire : les annuaires généralistes, les annuaires sélectifs et les annuaires et portails thématiques.

- Les annuaires généralistes sont les plus connus et recensent les sites grand public et professionnels, dans tous les domaines.
- Les annuaires sélectifs cherchent à avoir la couverture la plus complète possible du Web. Plus anciens que les répertoires généralistes, ces annuaires ont été réalisés par des professionnels de l'information, bibliothécaires ou documentalistes, qui sélectionnent uniquement les sites les plus à même de répondre aux diverses questions des chercheurs, des enseignants et des étudiants. Ces annuaires couvrent le plus souvent toutes les matières enseignées dans les universités, et ne référencent dans chaque catégorie que les sites les plus riches[7].
- Les annuaires et portails thématiques sont construits sur le même principe que les annuaires généralistes, mais ont une approche thématique : leur objectif est de recenser les ressources dans un domaine précis. Les mêmes raisons qui ont conduit de nombreux annuaires généralistes à se transformer en portails ont fait évoluer nombre d'annuaires thématiques vers le portail thématique appelé parfois « vortail[8] ».

- Les moteurs de recherche

La vocation des moteurs de recherche, apparus sur le Web dans les années 1990, est d'indexer le plus grand nombre de pages Web. Ces moteurs sont accessibles à l'aide de requêtes par mots-clés sur le texte intégral des pages.

Trois éléments composent les moteurs de recherche : un robot, un index, un serveur Web.

- Le robot, également appelé « *crawler* » ou « *spider* », parcourt le Web de façon automatique, de lien en lien. Le robot de Google, par exemple, s'appelle Googlebot. En partant d'un certain nombre d'adresses définies au préalable, le robot va tester tous les liens hypertextes qu'il rencontre et rapatrier le contenu des pages dans sa base. Il va également visiter toutes les pages soumises par les éditeurs de sites, dans le cadre du référencement. Ce parcours de lien en lien à travers le Web lui demande en moyenne quatre semaines, ce délai pouvant varier de une à six semaines. Une fois le « tour du Web » achevé, le robot repart de sa base pour la mettre à jour et pour identifier de nouvelles pages.
- L'index contient tous les mots de toutes les pages rapatriées par le robot. Le plus souvent, le texte intégral de la page ainsi que ses différentes balises Méta[9] (titre, mots-clés, etc.) sont indexés, mais il peut y avoir des variantes selon les moteurs.
- Le serveur Web offre l'interface de recherche à l'utilisateur. Celle-ci lui permet de lancer une requête par mots sur l'index du moteur, avec des possibilités plus ou moins sophistiquées.

- Les métamoteurs

C'est à partir de ces outils que l'on peut évoquer le début d'une veille digne de ce nom. Comme aucun outil de recherche n'assure une couverture complète du réseau Internet, il est indispensable d'en interroger plusieurs si l'on souhaite avoir un panorama de ce qui existe sur le Net sur un sujet, ou tout simplement pour augmenter ses chances d'identifier des pages pertinentes.

C'est principalement le rôle des métamoteurs, désignés aussi sous le nom de « métamoteurs *on line* », que d'interroger plusieurs outils, successivement ou simultanément selon les cas.

Certains métamoteurs sont des logiciels qu'il faut télécharger préalablement et qui se connectent à Internet le temps de la recherche. Ils entrent dans la catégorie des « agents pour la recherche », mais fonctionnent en fait comme un métamoteur *off line*.

Parmi les nombreux métamoteurs existants, l'*Open Directory* en recense plus de 185 dans sa catégorie « *Metasearch Tools* ». Les plus sophistiqués enregistrent la requête de l'internaute, l'envoient simultanément à différents moteurs et annuaires (que l'on peut choisir dans une liste), puis rapatrient les réponses, dédoublonnent les résultats, les classent (par pertinence, thème) et offrent parfois la possibilité de vérifier la validité des liens.

- Les agents (intelligents ou non) de recherche d'information

Ils sont classés dans la catégorie des « agents pour la recherche » et sont parfois aussi appelés « métamoteurs *off line* ».

Parmi ces agents intelligents, on distingue deux catégories dont la première est celle des agents de recherche d'information qui sont des logiciels intermédiaires entre les moteurs de recherche et les agents intelligents. Même s'ils ne possèdent pas toutes les caractéristiques des agents intelligents, ils sont, pour la plupart, beaucoup plus évolués que de simples moteurs de recherche. D'ailleurs, de nombreux efforts sont déployés par les éditeurs de logiciels pour se rapprocher des agents intelligents.

Un agent intelligent doit pouvoir prendre des initiatives et agir sans intervention de

l'utilisateur final. Dans le contexte du Web, il doit pouvoir agir alors que l'utilisateur est déconnecté. Les éditeurs de logiciels contournent le problème de l'autonomie en programmant leurs logiciels afin qu'ils puissent, à intervalles réguliers, se connecter automatiquement au Web pour y effectuer les tâches confiées par les utilisateurs.

Dans ce cas-là, on ne peut véritablement parler d'autonomie car l'agent n'est pas actif en permanence et ne fait que reproduire comme un automate la requête programmée par l'utilisateur.

Les agents de recherche d'information sont d'une grande diversité et ils remplissent rarement tous les mêmes tâches. Cependant, on peut identifier cinq fonctionnalités principales :

1. la recherche ponctuelle d'information
   Il s'agit de pouvoir effectuer des recherches sur un thème précis. Cette recherche peut être plus ou moins approfondie et complète.
2. la mise en place d'une veille thématique
   Cette fonctionnalité permet à l'utilisateur de suivre en permanence l'évolution de l'information sur des thèmes qu'il aura lui-même définis. Le logiciel se charge de rechercher de manière automatique et périodique l'information souhaitée.
3. la gestion de l'information récoltée
   Il s'agit d'éditer, d'archiver, de détruire et de mettre à jour les résultats des recherches.
4. l'analyse des documents rapatriés
   Cette analyse va, par exemple, permettre de mettre en relation les différents documents, de manière à faire ressortir de la masse l'information pertinente. Il s'agit également de créer des résumés automatiques.
5. la navigation *off line*
   Cette fonctionnalité permet de naviguer parmi les informations trouvées sur Internet par le système sans nécessairement être connecté.

Il est important de noter que ces fonctionnalités peuvent être présentes dans de nombreux logiciels sans qu'elles mettent en jeu pour autant les mêmes technologies.

Pour la seconde catégorie, celle des agents intelligents, la recherche ne se fait pas depuis un site Web. Ces agents sont en effet des logiciels qui doivent être téléchargés et installés sur le disque dur de l'utilisateur dans sa version définitive ou dans sa version d'évaluation.

Dans une étude de Gilles Balmisse (2002) sur les agents, on en trouve la définition suivante :

> « On appelle agent intelligent une entité réelle ou abstraite qui est capable d'agir sur elle-même et sur son environnement, qui dispose d'une représentation partielle de cet environnement, qui, dans un univers multi-agent, peut communiquer avec d'autres agents et dont le comportement est la conséquence de ses observations, de sa connaissance et des interactions avec les autres agents. »
> (Voir : http://www.gillesbalmisse.com/IMG/pdf/GB_Agent.pdf, page 3)

Si le principe de fonctionnement des agents est le même que celui des métamoteurs sur le Web, leurs possibilités plus étendues et en constante évolution en font toutefois des outils de recherche plus performants.

En dehors du téléchargement des résultats qui pourront être enregistrés pour un examen ultérieur, les agents intelligents proposent de plus en plus d'éditer un rapport de recherche reprenant la liste des résultats, sous la forme d'une page HTML.

Enfin, et surtout en ce qui nous concerne, ces agents intelligents se transforment de plus en plus en véritables automates de veille que l'on peut paramétrer pour se déconnecter une fois le travail terminé, envoyer un rapport de recherche signalant les nouvelles pages, télécharger des documents sur le disque, éliminer les liens non valides, etc.

Pour asseoir cette découverte des outils de veille sur Internet, l'auteur de cet article vous recommande la lecture de l'ouvrage de Samier et Sandoval (2002), « La veille stratégique sur l'Internet ».

Conclusion : vers une écologie informationnelle

L'espace informationnel est devenu complexe et le devient chaque jour davantage.

À la diversité des documents dans leur nature, leur forme et leur intention de communication, s'ajoute aujourd'hui une diversité des accès technologiques et des circuits de production et de diffusion. Les nouvelles mises en scène de l'information, induites par des changements de supports, impliquent cependant de nouveaux modes de lecture et de prise d'information dont l'impact fait encore l'objet de recherches. Pour l'apprenant comme pour l'enseignant, la recherche d'information en vue d'acquérir une connaissance reste à ce jour une aventure où chacun part à la découverte de l'inconnu, et il faut donc que le voyage soit tentant.

L'accès à l'information et à des documents implique encore plus qu'hier une écologie informationnelle nécessitant une compréhension de leurs modes d'élaboration et de diffusion.

La diversité des outils décrits dans cet article montre que les techniques électroniques de traitement de l'information et de communication jouent les rôles les plus divers dans la formation à l'activité intellectuelle. Dans ce contexte, l'ensei-

gnant, s'il veut rester au cœur de l'organisation et de la diffusion des savoirs, doit s'efforcer de montrer à ses étudiants qu'un même message peut être véhiculé par diverses technologies dont il se doit de maîtriser les composantes pour caractériser l'espace de ressources dans lequel se situe la classe, le groupe ou encore le département d'université.

La base de partage de connaissances et de veille « Commun@utice » (http://communautice.icl-lille.fr/veille.php), que nous avons réalisée, synthétise au sein d'une mallette de veille une sélection de liens, de ressources et d'outils déposés par près de 600 étudiants dans le cadre d'exercices de veille et de mises en situation diverses validés par les experts (enseignants de différentes matières) sur presque deux années de veille.

Une nouvelle version ouvrira ses portes en mai 2005 et verra naître le concept de « salles de veille » où apprenants comme enseignants contribueront à l'élaboration du savoir en vue d'apprendre à apprendre ensemble.

L'objectif final consiste à aider tous les acteurs de l'éducation à faire leurs premiers pas dans le monde de la veille et de l'intelligence pédagogiques.

Références

Achart, P. et Bernat, J.-P. (1998). *L'intelligence économique, mode d'emploi*. Paris : ADBS.

Gauthier, M. (1998). *Pays de connaissances*. Monaco : Du Rocher.

Balmisse, J. (2002). *Accès à la connaissance, outils et applications pratiques*. Paris : Vuibert.

Barbot, M. J. et Camatarri, G. (1999). *Autonomie et apprentissage, l'innovation dans la formation*. Paris : Presses universitaires de France.

Bulinge, F. (2002). *Pour une culture de l'information dans les petites et moyennes organisations : un modèle incrémental d'Intelligence économique*, Université de Toulon et du Var, France.

Cros, R. C., Gardin, J.-C. et Lévy, F. (1964). *L'automatisation des recherches documentaires. Un modèle général*. Le Syntol. Paris : Gauthier-Villars.

Dou, H. (1995). *Veille technologique et compétitivité. L'intelligence économique au service du développement industriel*, Paris : Dunod.

Foenix-Riou, B. (2002). *Guide de recherche sur Internet, outils et méthodes*. Paris : ADBS, Nathan.

Foenix-Riou, B. (2001). *Recherche et veille sur le Web visible et invisible*. Paris : Tech et Doc.

Guir, R. (2002). *Pratiquer les TICE, former les enseignants et les formateurs à de nouveaux usages*. Bruxelles : De Boeck.

Lagane, C. (2004, 21 juin). Les moteurs de recherche, portes d'entrée du Web. Site CRN-VNUnet, section *Actualités*. Récupéré le 6 juin 2005 de http://www.vnunet.fr/actualite/logiciels/utilitaires/20040621006

Moeglin, P. (1996). Multimédias et éducation, le démon de la convergence. Dans *Outils multimédias et stratégies d'apprentissage du FLE*, Cahiers de la Maison de la Recherche, Lille III, Université Charles-de-Gaulle.

Morizio, C. (2002). *La recherche d'information*. Paris : ADBS, Nathan.

Norman, D. A. (1991). *Cognitive artifacts in designing interaction: Psychology at the human-computer interface*. New York : Cambridge University Press.

Perriault, J. (2002). *Éducation et nouvelles technologies, théorie et pratiques*. Paris : Nathan.

Perriault, J. (2002). *L'accès au savoir en ligne*. Paris : Odile Jacob.

Pinte, J.-P. (2004), *Le rôle de la veille pédagogique*, 21e Congrès AIPU, Les technologies éducatives dans l'enseignement supérieur. Marrakech, du 3 au 7 mai 2004.

Prax, J.-Y. (2003). *Le Manuel du knowledge management, une approche de 2ème génération*. Paris : Dunod.

Revelli, C. (2000). *Intelligence stratégique sur Internet – Comment développer des activités de veille et d'intelligence économique sur le Web?* Éd. Paris : Dunod.

Samier, H. et Sandoval, V. (2002). *La veille stratégique sur l'internet*. Paris : Hermes Science.

UDC Consortium (n. d.). *La Classification Décimale Universelle (C.D.U.) et le monde francophone*. Site Universal Decimal Classification [UDC] Consortium. Récupéré le 6 juin 2005 de http://www.udcc.org/cdu.htm

Notes

[1] Jean-Yves Prax définit la sérendipité comme l'art de trouver ce qu'on ne cherche pas (2003, p. 76).

[2] Il y a infobésité lorsque la quantité d'informations disponible est si importante qu'il devient impossible pour les individus ou les organisations d'en extraire rapidement et efficacement l'information pertinente. (En anglais : *information overload.*) Voir le Schéma 1.

[3] Il ne s'agit pas d'une autonomie positive, dans ce cas, mais de la capacité à s'adapter à un matériel d'apprentissage.

[4] Le « Web visible » est l'ensemble des pages Web qui peuvent être indexées par les moteurs de recherche, par opposition aux pages du « Web invisible », que les robots des moteurs ne peuvent atteindre.

[5] On appelle requête toute question posée à un système documentaire.

[6] Pour veiller à l'évolution des outils de recherche sur le Web et découvrir leur fonctionnement, il convient de découvrir le site référence en ce domaine : http://outils.abondance.com/
Une description des principaux annuaires sélectifs est donnée dans la rubrique du même nom sur le site de la Bibliothèque Nationale de France (http://www.bnf.fr/pages/liens/).

[8] Francisation du terme anglais *vortal*, employé pour *vertical portal*.

[9] Une balise Méta est une information insérée par le webmestre qui ne s'affiche pas à l'écran et qui donne des indications sur la page (titre, mots-clés).